



TEAMS

Technical Evaluation Assistance in Mathematics & Science



EXTERNAL FEEDBACK FOR EDUCATIONAL RESEARCH: WHAT SHOULD WE ASK?

PREPARED FOR:
THE NATIONAL SCIENCE FOUNDATION
DIRECTORATE FOR EDUCATION AND HUMAN RESOURCES
MATH AND SCIENCE PARTNERSHIP RESEARCH AND TECHNICAL ASSISTANCE PROGRAM

PREPARED BY:
RMC RESEARCH CORPORATION
633 17TH ST., SUITE 2100
DENVER, CO 80202

APRIL 2016



TEAMS

Technical Evaluation Assistance in Mathematics & Science



EXTERNAL FEEDBACK FOR EDUCATIONAL RESEARCH: WHAT SHOULD WE ASK?

PREPARED UNDER GRANT:
DRL-1238120

PREPARED BY:

JOHN T. SUTTON
ARLENE MITCHELL
CATHERINE CALLOW-HEUSSER
MICHAEL J. CULBERTSON
EMMA ESPEL
JENNIFER WESTON-SEMENTELLI

RMC RESEARCH CORPORATION
633 17TH STREET, SUITE 2100
DENVER, CO 80202

PREPARED FOR:
THE NATIONAL SCIENCE FOUNDATION
DIRECTORATE FOR EDUCATION AND HUMAN RESOURCES
MATH AND SCIENCE PARTNERSHIP RESEARCH AND TECHNICAL ASSISTANCE PROGRAM

APRIL 2016

NOTE: Any opinions, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ACKNOWLEDGMENTS

FOR QUESTIONS ABOUT THIS REPORT, PLEASE CONTACT JOHN SUTTON
AT 1-800-922-3636, OR E-MAIL AT SUTTON@RMCRES.COM.

RMC RESEARCH CORPORATION IS AN EQUAL EMPLOYMENT OPPORTUNITY
AND AFFIRMATIVE ACTION EMPLOYER AND A DRUG-FREE WORKPLACE.

TABLE OF CONTENTS

SITUATING THE WORK	1
PROJECT DESCRIPTION	1
PURPOSE FOR THIS PUBLICATION	1
INTRODUCTION	1
RESEARCH AND EVALUATION OF RESEARCH	3
RELATIONSHIP BETWEEN RESEARCH TYPES, RESEARCH PHASES, AND EXTERNAL FEEDBACK MODELS	5
SUMMARY OF RESEARCH TYPES FROM THE COMMON GUIDELINES	5
RESEARCH PHASES: DESIGN, IMPLEMENTATION, OUTPUT, AND ACCOUNTABILITY	8
EXTERNAL FEEDBACK MODELS FOR EVALUATION OF RESEARCH	9
PEER REVIEW OF THE RESEARCH PROPOSAL	9
AGENCY MONITORING AND REVIEW	11
EXTERNAL REVIEW PANELS OR/AND ADVISORY BOARDS	11
THIRD-PARTY OR EXTERNAL EVALUATOR	12
PEER REVIEW FOR PUBLICATIONS AND CONFERENCE PRESENTATIONS	12
PUTTING IT ALL TOGETHER	13
PLANNING FOR EVALUATION OF RESEARCH	14
USE OF MULTIPLE EXTERNAL FEEDBACK MODELS	14
EVALUATION QUESTIONS THAT SUPPORT EVALUATION OF RESEARCH	15
STRENGTHENING EVALUATION PLANS IN RESEARCH PROPOSALS	17
CONCLUSION	18
REFERENCES AND ARTICLES REVIEWED TO DEVELOP MODEL FOR EVALUATION OF RESEARCH	19
APPENDIX	A-1

TABLE OF EXHIBITS

EXHIBIT 1.	EVALUATION OF RESEARCH SITUATED IN EDUCATIONAL INTERVENTION	4
EXHIBIT 2.	RESEARCH TYPES, RESEARCH PHASES, AND EXTERNAL FEEDBACK MODELS	5
EXHIBIT 3.	DESCRIPTION AND PURPOSE OF RESEARCH TYPES	6
EXHIBIT 4.	DESCRIPTION OF PHASES OF THE RESEARCH CYCLE	8
EXHIBIT 5.	DESCRIPTION OF EXTERNAL FEEDBACK MODELS	9
EXHIBIT 6.	DIFFERENT USES OF PEER REVIEW	10
EXHIBIT 7.	RELATIONSHIP OF RESEARCH TYPES, PHASES OF THE RESEARCH CYCLE, AND EXTERNAL FEEDBACK MODELS	13
EXHIBIT 8.	EXTERNAL FEEDBACK MODELS ALIGNED TO RESEARCH TYPES	15
EXHIBIT 9.	SAMPLE QUESTIONS BY PHASE	16

APPENDIX EXHIBITS

EXHIBIT A1.	PRE-FUNDING AND DESIGN EVALUATION QUESTIONS	A-1
EXHIBIT A2.	IMPLEMENTATION EVALUATION QUESTIONS	A-4
EXHIBIT A3.	OUTPUTS EVALUATION QUESTIONS	A-5
EXHIBIT A4.	ACCOUNTABILITY EVALUATION QUESTIONS	A-6
EXHIBIT A5.	COMMON GUIDELINES EVALUATION QUESTIONS	A-7

SITUATING THE WORK

PROJECT DESCRIPTION

The Technical Evaluation Assistance in Mathematics and Science (TEAMS) project (DRL#1238120) is funded under the National Science Foundation (NSF) Mathematics and Science Partnership (MSP) Research Evaluation and Technical Assistance (RETA) program. TEAMS began operations in May 2013 with the specific goal to: *Strengthen the quality of MSP project evaluation and build the capacity of evaluators by strengthening their skills related to evaluation design, methodology, analysis, and reporting.* Any opinions, suggestions, and conclusions or recommendations expressed in this paper are those of the TEAMS project and do not necessarily reflect the views of the NSF; NSF has not approved or endorsed its content.

PURPOSE FOR THIS PUBLICATION

This paper, prepared by TEAMS staff, provides insights into the evaluation of educational research in order to strengthen research projects and increase the quality of evidence about what works to improve student learning from the lens of science, technology, engineering, and mathematics (STEM) education. In particular, we seek to increase the usability of research findings through evaluation approaches designed specifically to improve the processes and results of research in STEM education. The primary questions that guided the development of this paper are:

- *What are the defining features of evaluation of research in contrast with both educational research and program evaluation?*
- *What is the relationship between research types, phases of the research cycle, and evaluation feedback models?*
- *What evaluation feedback models are best suited to different research types?*
- *What evaluation questions guide the evaluation of research projects?*
- *How should evaluation questions and feedback models be adapted for each research type as defined by the Common Guidelines for Educational Research and Development?*

INTRODUCTION

In August 2013, the NSF and the Institute for Education Sciences (IES) jointly released the *Common Guidelines for Educational Research and Development* (<http://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf>), which established a typology for educational research and provided corresponding guidelines for “improving the quality, coherence, and pace of knowledge development in science, technology, engineering and mathematics (STEM) education.” In particular, the recommendations were developed to help ensure that research is designed to improve the quality of evidence about what works to improve student learning and to gain greater return on investments in research.

The *Common Guidelines* indicate that research of all types should incorporate external feedback, and the document provides the following guidelines for this feedback:

[A research] project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator, and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings (p. 23).

However, resources to guide the form and content of external feedback for educational research are limited. Work in research evaluation is well underway in the medical field (Guthrie, Wamae, Diepeveen, & Grant, 2013) and across disciplines in higher education institutions internationally (Derrick & Pavone, 2013; Research Excellence Framework, 2012). However, there appears to be a lack of scholarly work in research evaluation that focuses on educational initiatives, such as STEM projects and programs. Moreover, Guthrie et al. (2013) suggest that traditional evaluation tools may be wholly unsuitable for evaluating research and suggest that new tools and frameworks are needed.

Given the current limitations regarding recommendations, models, or tools for evaluating research, the TEAMS project conducted a systematic literature review in fall 2015.¹ A review of more than 50 published articles was conducted to provide insights into how different feedback models would be used to evaluate research. This review provided a limited yield of direct examples of when and how specific feedback models would be most useful or appropriate for evaluating different types or stages of research. Based, in part, on this literature review, the TEAMS project concurs with the suggestion from the RAND Corporation (Guthrie et al., 2013). that “new tools and frameworks are needed” in the evaluation of educational research as well.

This paper articulates a range of illustrative evaluation questions that may guide the design of external feedback plans for educational research. The evaluation questions are organized with attention to their appropriateness for research of different types (according to the typology from the *Common Guidelines*), various feedback models (described below), and the phases of a research project (design, implementation, outputs, and accountability). In addition, these questions can help guide the proposal preparation process to ensure consideration regarding the selection and application of the best feedback model and research evaluation questions for any research study. This paper contributes to knowledge in the area of educational research evaluation and provides a resource for answering

¹ The literature review used Google Scholar, PsycINFO, and ERIC to identify articles with specified keywords that addressed the topic of the paper and that were published within the last 10 years. Other relevant resources were identified as appropriate during review of general searches, scanning reference lists, and in discussion with national experts. Four independent reviewers decided whether or not initial resources would be included in the final search. Selection criteria included publicly available, no-cost documents that identify, articulate or convey evaluation models and research types in education (primarily), science education, and/or mathematics education with the exception of Masters or Doctoral dissertations, which were not included. Discrepancies were discussed and resolved. A final list of more than 50 published articles was included in the literature summary. These articles are included in the reference list.

questions of the quality of STEM research projects at the local level and beyond—for learning, improvement, and accountability.

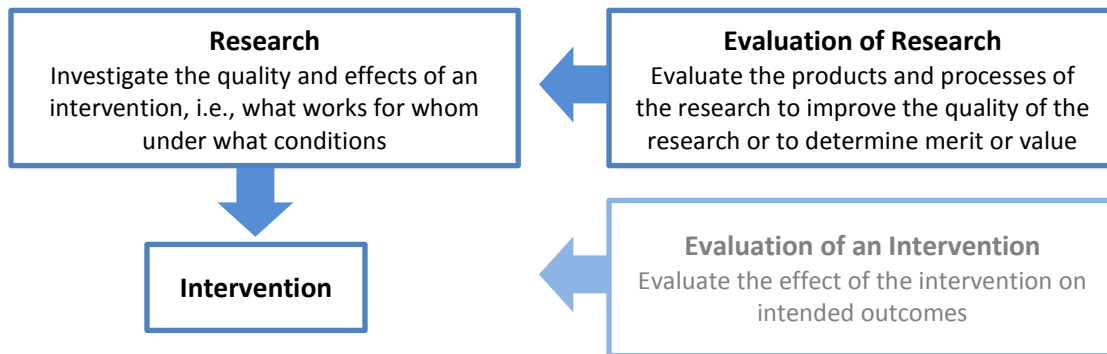
RESEARCH AND EVALUATION OF RESEARCH

In order to better understand the evaluation of educational research, it is important to differentiate between research and the evaluation of research. At first glance, research and evaluation can look very similar. They both involve systematic inquiry, and many times research and evaluation studies will use very similar methods for data collection and analysis (particularly for research in the social sciences). Recognizing these similarities, the literature review showed that evaluation is sometimes referred to as *evaluation research*, indicating that evaluation is a particular type of research (that is, a particular type of systematic inquiry). And yet, evaluation is somewhat distinct from research. While the distinctions could be nuanced in a number of different ways, two important distinctions are related to values and locality. All research is necessarily value laden (at a bare minimum, the very act of engaging in research makes a statement about the value of a particular research question). However, evaluation places values at the forefront of investigation: the systematic inquiry conducted by evaluators is precisely about the quality or condition of the evaluand (that is, the thing being evaluated) relative to some system of values. Some evaluations end with a summative statement about the evaluand's merit or worth, while other evaluations describe strengths or weaknesses for improvement or needs assessment purposes. In distinction, research generally places less explicit emphasis on the value of or values associated with the subject of inquiry, favoring instead descriptions of other aspects of the subject.

Second, evaluations are typically local in scope—not necessarily local *geographically* (evaluations can certainly be national or international in geography), but local to the particular evaluand under study. The goal of an evaluation is to understand *this* program, *this* policy, *this* organization, *this* candidate, *this* research project, and so forth. Research, on the other hand, is generally interested less in the particular subjects or participants in the given study and more with how the characteristics of the given subjects can lead to knowledge about the world that generalizes *beyond* those subjects, although case study and some forms of qualitative constructivist research may be exceptions to this tendency. These two characteristics—value and locality—are helpful for distinguishing between research activities and evaluation activities in the evaluation *of* research. In a project that contains both research activities and the evaluation of those research activities, the research activities consist of inquiry intended to generalize beyond the particular project. Evaluation activities, on the other hand, consist of inquiry local to the particular project to describe the quality or other aspects of value of those research activities. The evaluation activities provide formative feedback to improve project processes and products, or summative feedback that describes the project's merit or value.

In educational research, the focus of the research is often on the quality of an *intervention*. As such, the research is situated in the *context* of the educational intervention. Evaluation of research, on the other hand, is typically focused on the quality of the *research* and is not an evaluation of the quality of the intervention. Rather, the evaluation of research indicated in the *Common Guidelines* refers to the quality of the *research* that is potentially situated in some educational intervention, as conveyed through Exhibit 1.

EXHIBIT 1. EVALUATION OF RESEARCH SITUATED IN EDUCATIONAL INTERVENTION



This distinction is particularly important for design and impact studies in which the research objectives are very closely tied with understanding the quality and effects of interventions. In such research, even though the research questions may have to do with the quality of an intervention, the evaluation activities intended to satisfy the external feedback models in the *Common Guidelines* are about the *process and products of the systematic inquiry* on the intervention, and not about the intervention itself—as such, evaluation of research in these settings could be considered a form of meta-evaluation.

RELATIONSHIP BETWEEN RESEARCH TYPES, RESEARCH PHASES, AND EXTERNAL FEEDBACK MODELS

To gain insights into the relationship between research types, phases of the research cycle, and the evaluation feedback models, as shown in the following exhibit, we will begin by providing descriptions of each. We will then present a model under which the research types, phases of the research, and feedback models all come together to support the generation of knowledge.

EXHIBIT 2. RESEARCH TYPES, RESEARCH PHASES, AND EXTERNAL FEEDBACK MODELS

Research Types	Research Phases	External Feedback Models
Foundational	Design	Agency Monitoring
Early Stage/Exploratory	Implementation	External Evaluator
Design/Development	Output	External Review Panel/ Advisory Board
Efficacy	Accountability	Peer Review
Effectiveness		
Scale-up		

SUMMARY OF RESEARCH TYPES FROM THE COMMON GUIDELINES

The *Common Guidelines* delineate six basic research types (Exhibit 3): Foundational, Early-Stage or Exploratory, Design and Development, Efficacy, Effectiveness, and Scale-up research. The first two types of research, Foundational and Early-Stage or Exploratory research, contribute to core knowledge in education including the basic understanding of teaching and learning. Design and Development research works to develop solutions to achieve a goal. The last three types of research, Efficacy, Effectiveness, and Scale-up research, all contribute to the evidence of impact to generate reliable estimates of the ability of the strategy or intervention to achieve its goal. Each research type has nuanced characteristics that are pertinent to decisions in evaluation feedback model choice and evaluation question development. Although this research typology reflects a gradation in the types of evidence that a research program could generate over time, it is important to remember that research is not linear—research programs may skip steps or loop back to earlier stages, and any given research project may tackle features of multiple types of research at once.

EXHIBIT 3. DESCRIPTION AND PURPOSE OF RESEARCH TYPES

Research Type	Description	Purpose
#1: Foundational Research	Provides the fundamental knowledge that may contribute to improved learning and other relevant education outcomes. Studies of this type seek to test, develop, or refine theories of teaching or learning and may develop innovations in methodologies and/or technologies that will influence and inform research and development in different contexts.	The purpose of Foundational research is to advance the frontiers of education and learning; develop and refine theory and methodology, and provide fundamental knowledge about teaching and/or learning. Foundational research studies may examine phenomena without establishing an explicit link to education outcomes.
#2: Early-Stage or Exploratory Research	Examines relationships among important constructs in education and learning to establish logical connections that may form the basis for future interventions or strategies to improve education outcomes. These connections are usually correlational rather than causal.	The purpose of Early-Stage or Exploratory research is to investigate approaches to education problems to establish the basis for design and development of new interventions or strategies, and/or to provide evidence for whether an established intervention or strategy is ready to be tested in an efficacy study. Early-Stage or Exploratory research should establish initial connections to outcomes of interest. Studies in this genre should support the development of a well-explicated theory of action that can inform the development, modification, or evaluation of an intervention or strategy.
#3: Design and Development Research	Draws on existing theory and evidence to design and iteratively develop interventions or strategies, including testing individual components to provide feedback in the development process.	<p>The purpose of Design and Development research is to develop new or improved interventions or strategies to achieve well-specified learning goals or objectives, including making refinements on the bases of small-scale testing. Typically this research involves four components:</p> <p>(1) Develop a solution based on a well-specified theory of action appropriate to a well-defined end user; (2) Create measures to assess the implementation of the solutions; (3) Collect data on the feasibility of implementing the solution(s) in typical delivery settings by intended users; and (4) Conduct a pilot study to examine the promise of generating the intended outcomes.</p>

Research Type	Description	Purpose
#4: Efficacy Research	Allows for testing of a strategy or intervention under “ideal” circumstances, including with a higher level of support or developer involvement than would be the case under more rigorous research.	<p>The purpose of Efficacy research is to determine whether an intervention or strategy can improve outcomes under what are sometimes called “ideal” conditions. For example, these conditions may include more implementation support of more highly trained personnel that would be expected under routine practice, or in contexts that include a more homogeneous sample of students, teachers, schools, and/or districts than is typical.</p> <p>Efficacy studies may involve the developer in the implementation of the intervention or strategy; however, the study should include reasonable safeguards for ensuring the objectivity and integrity of the study. Sometimes Efficacy studies are used to replicate previous evaluations of an intervention, but under different conditions.</p>
#5: Effectiveness Research	Examines effectiveness of a strategy or intervention under circumstances that would typically prevail in the target context.	The purpose of Effectiveness research is to estimate the impacts of an intervention or strategy when implemented under conditions of routine practice. To this end, implementation should be similar to what would occur if a study were not being conducted. An Effectiveness study should be carried out with no more developer involvement than what would be expected under typical implementation.
#6: Scale-up Research	Examines effectiveness in a wide range of populations, contexts, and circumstances, without substantial developer involvement in implementation or evaluation.	The purpose of Scale-up research is to estimate the impacts of an intervention or strategy under conditions of routine practice and across a broad spectrum of populations and settings. That is, Scale-up research studies should be conducted in settings and with population groups that are sufficiently diverse to broadly generalize findings. As with Effectiveness research, Scale-up research should be carried out with no more developer involvement than what would be expected under typical implementation.

RESEARCH PHASES: DESIGN, IMPLEMENTATION, OUTPUT, AND ACCOUNTABILITY

The research evaluation inferred in the *Common Guidelines* is specific to research regarding interventions to improve some outcome. The language of “reviews of [the research] design and activities,” “ongoing monitoring,” and intending to “influence the project’s activities and improve the quality of its findings” (p. 23) suggests that the particular type of research evaluation envisioned by the *Common Guidelines* is primarily formative in nature, closely related with the process and immediate outputs of the research. A framework for evaluation of MSP projects identified the Design-Implementation-Outcomes (DIO) Cycle of Evidence (Callow-Heusser, Chapman, & Torres, 2005) which could be appropriately applied to any feedback model. As such, in this paper, we consider research evaluation that covers three phases of the research cycle similar to the DIO: *design*, *implementation*, and *outputs* (Exhibit 4). We use the term outputs, rather than outcomes, because we are referring to the direct result of the research activities, and not to how the implementation of an intervention that is being investigated or research activities affect stakeholders or systems. To this, we add evaluation that pertains to *accountability* of the research grantees provide to their funder insofar as the accountability is related to the design, implementation, and outputs of the research project.

EXHIBIT 4. DESCRIPTION OF PHASES OF THE RESEARCH CYCLE

Research Phases	Description
Design	In the design phase, researchers determine what they will investigate and how they will answer the research questions.
Implementation	The implementation phase consists of carrying out the research activities.
Outputs	The output phase consists of the development and dissemination of products based on the research, such as program materials, research articles, and presentations.
Accountability	The accountability phase consists of demonstrating that funding and other resources were used efficiently and effectively.

(Callow-Heusser et. al., 2005; Guthrie et al., 2013; Milessi et al., 2014.)

Evaluation can also take place throughout a number of different phases of the research cycle. During the proposal phase, research plans are evaluated for rigor, merit, and alignment with funders’ goals. It is quite common for many institutions/agencies to use a pre-funding design phase to help frame and inform the research design. Peer-reviewed publication venues make similar evaluations of research quality on the back end based on presentation of the results. Literature reviews examine the conclusions of research studies in the context of a growing body of knowledge on a given topic. The impact of research on a given field or on society may be considered in the evaluation of research funding programs or, in the case of employment in tenured positions, in personnel decisions about particular researchers. While all of this constitutes research evaluation, these forms of research evaluation generally fall outside the scope of activities intended by the external feedback models in the *Common Guidelines*.

EXTERNAL FEEDBACK MODELS FOR EVALUATION OF RESEARCH

The external feedback provided about research projects can take many forms. The *Common Guidelines* list peer review, ongoing monitoring and review by funding agency personnel, external review panels or advisory boards, and external evaluators (Exhibit 5). The aims of the evaluation process should be decided a priori and guide the development of the evaluation approach, including the choice of feedback model (Guthrie et al., 2013).

EXHIBIT 5. DESCRIPTION OF EXTERNAL FEEDBACK MODELS

External Feedback Model	Description
Peer review of the proposed project.	Peer review panels are operated by the funding agency. After submission of a proposal to a funding agency peer reviewers are assigned to a project to assess the merit of the application for funding. Both IES and NSF provide feedback to Principal Investigators on grant submissions.
Ongoing monitoring and review by the grant making agency's personnel.	Monitoring and review can take multiple forms. NSF and IES require a yearly report to update the agency on the progress of the grant and the budgeting of funds. Some grants may be subject to a more complete audit of procedures, expenditures, or both.
External review panels or advisory boards proposed by the project and/or the agency.	External review panels or advisory boards are typically a group of individuals with specific expertise assembled by the project. The review panels or advisory boards are often managed by key project personnel, an evaluator, or researcher and represent the relevant stakeholders. The members of the review panels or advisory board are knowledgeable in areas such as specific content, evaluation, research methodology, and/ or location of research project. Advisory Boards can vary in size but should include individuals with varied expertise who can fulfill the roles and responsibilities envisioned by the project.
A third-party evaluator.	A third-party evaluator is an independent third-party responsible for assessing the implementation of project activities. Additionally the third-party evaluator engaged in documenting the successes and difficulties of a project in relation to the initial project goals and mid-course corrections.
Peer review of publications and conference presentations resulting from the project.	Peer review of publications and conference presentations provide feedback to the project. The feedback is typically provided by reviewers following submission of research findings to conferences and journals. Reviews are generally anonymous and completed by individuals with expertise in the content area and/or methodology.

PEER REVIEW OF THE RESEARCH PROPOSAL

While peer review is typically based on the outputs of projects, as suggested in the previous table, peer review used in an external feedback model can take at least three forms, as presented in Exhibit 6.

EXHIBIT 6. DIFFERENT USES OF PEER REVIEW

Application	Peer Review Feedback Model
During Proposal Development	Prior to submission of a proposal and while an institution or organization is developing a proposal for consideration by a funding agency, the Principal Investigator, Leadership or Development Team, or Office of Sponsored Programs may pull together a group of reviewers. The reviewers are expected to examine the proposal and offer suggestions or recommendations as a process to determine the fundability of the proposal and increase the likelihood of the proposal being funded. Those individuals or groups involved as peer reviewers at this stage can work as volunteers or as members of the partner institutions, although often they are hired as consultants by an institution or organization.
Determination For Funding	Many funding agencies (e.g., NSF, IES, National Institutes of Health [NIH], National Aeronautics and Space Agency [NASA], Foundations, State Education Agencies, etc.) use a peer review panel as a mechanism or process for making initial funding decisions. These review panels are usually comprised of a small number of reviewers to facilitate open dialogue and discussion regarding the merits of the proposal(s) being reviewed. The reviewers involved in this peer review process typically possess expertise, knowledge, and/or experience related to at least one aspect or dimension of the solicited proposal. That expertise, knowledge, and/or experience can be related to the content or pedagogical content knowledge for the proposed discipline(s) for which the proposal is being submitted, research design, measurement, evaluation, professional development, assessment, leadership, technology, or some other important and relevant area. This peer review process typically applies some sort of rubric and/or standardized review process with results provided back to the proposer, whether funded or not. These reviews can take place in person or through some virtual process. Those individuals involved as peer reviewers at this stage typically receive a small stipend from the funding agency for their service.
Technical Working Group (TWG)	A project may choose to convene a TWG to focus on a particular topic, and to examine and discuss specific aspects of the project program, research, implementation, and evaluation designs and deliverables. A TWG is typically an ad hoc group of experts on a particular topic who work together on specific goals. These meetings are often meant to encourage TWG members to discuss the current state of research on a topic and/or to identify potential shortcomings in the project design. It is not uncommon for TWG members to review, edit, and revise various documents for consideration by the funding agency. Additionally, TWG members are often individually consulted by a member of the project around a specific feature of the project. The TWG can be convened in person, virtually, or through a combination of mechanisms. Individuals involved as members of a TWG typically receive a small stipend or consultant fee from the project for their services.

AGENCY MONITORING AND REVIEW

The manner in which a funding agency monitors the funded project can take on a variety of different structures. Additionally, the level of agency monitoring and review is frequently related to the type of award. In some instances, the monitoring is considered more of an audit and focused on the fiscal aspects of the funded research. This sort of monitoring, which would typically take place on a routine basis (e.g., monthly, quarterly, annually, etc.) is intended to ensure expenditures are in line with the funding organization's policies and practices, and in many instances, ensure that resources are being expended at levels consistent with proposed activities to minimize a large accumulation of unspent funds. Another facet of agency monitoring is undertaken by a single individual tasked with serving as the funding agency's liaison with the project. This role, often referred to as Program Officer or Project Liaison, interacts with the funded project to review the work being done and ensure that what was proposed occurs in a timely and appropriate manner. Additionally, the Program Officer can serve as a conduit between the funding agency and/or organization and the Principal Investigator, engaging in ongoing communications to inform the Principal Investigator about any changes in the funding agency's policies/practices, share information about other opportunities that are available, and help to solve problems the project may encounter around any aspect of the work. It is the role of the Program Officer to gather evidence from the project on a routine basis and maintain the funding agency's records regarding project staff changes, program changes or mid-course corrections, and annual reports. The person who serves in this capacity is typically an employee of the funding agency, or is contracted through any number of mechanisms to carry out the oversight role of the funding agency.

EXTERNAL REVIEW PANELS OR/AND ADVISORY BOARDS

An external review panel or advisory board is a committee or group without governing authority or responsibility for the project that is put together and managed by key project staff, evaluator, or researcher to provide technical advice, legitimacy, credibility, and /or prestige. Most often, the group is composed of individuals with evaluation, research, and content knowledge expertise and experience. Advisory boards serve a number of key functions, which include: (a) fostering stakeholder engagement; (b) maximizing external credibility; (c) facilitating political conciliation; (d) encouraging methodological integrity, and (e) promoting use of evaluation findings (Mattessich, 2012; Cohen, 2012). Methodological integrity may include planning and designing the evaluation, suggesting culturally sensitive protocols, identifying and focusing evaluation questions, defining outcomes and indicators, recommending data-collection methods and tools, providing advice and access to networks for engaging community-based staff and participants, and discussing and interpreting evaluation findings.

Mattessich (2012) suggested that an advisory board should include as many members as is necessary to adequately fulfill these key functions. The advisory board could be comprised of a variety of stakeholder groups, such as the funders' representatives, researchers and content experts, other resource controllers and regulators, persons affected by the program design, as well as thought leaders and critics. It may also include those with expertise and/or experience in the research being evaluated, the program, service, or its host organization, and other relevant aspects of a particular evaluation, type of evaluation, evaluation function, or evaluation unit (Baizerman, Fink, & Roholt, 2012). This varied composition serves to ensure the project is examined from multiple perspectives to promote rigor, responsiveness, and success. When considering candidates to serve on an external review panel or advisory board, the project should recognize preferred characteristics or traits for members. Some advisory board member characteristics that are important to consider include the following: not intimidated by the research or evaluation process; communicates and works well on a team; are willing to speak up, question, and challenge; connected to

relevant communities and cultural groups; interested in community and project well-being; are able to support project needs beyond personal interests; and contributes unique knowledge, skills, networks.

When forming an advisory board, the purpose and goals of the evaluation should be articulated prior to forming the board, as the project's goals for the board may drive member selection. In addition, the role of the board may drive decisions regarding frequency of meetings or board member compensation.

THIRD-PARTY OR EXTERNAL EVALUATOR

The third-party evaluator is an individual, organization, and/or agency that is not an employee of the funded organization that is carrying out the research, therefore making them “external” to the project. The intent of a third-party or external evaluator is to secure an objective view of the proposed research to help inform the project and report unbiased results from the evaluation. The external evaluator may be someone who has worked with people involved with the project, including the Principal Investigator, Co-Principal Investigator, researchers, and or other organization staff. The external evaluator may be someone who was recommended by a known associate or has a reputation for conducting external evaluations for similar research projects. The external evaluator usually has knowledge, expertise, and/or experience in a particular discipline or methodology that will allow them to carry out the external evaluation, or who will work with others in the evaluator's organization that brings that expertise to the evaluation. The external evaluator is usually identified in the submitted proposal and is funded through a sub-award and/or other contractual arrangement with the Principal Investigators organization.

PEER REVIEW FOR PUBLICATIONS AND CONFERENCE PRESENTATIONS

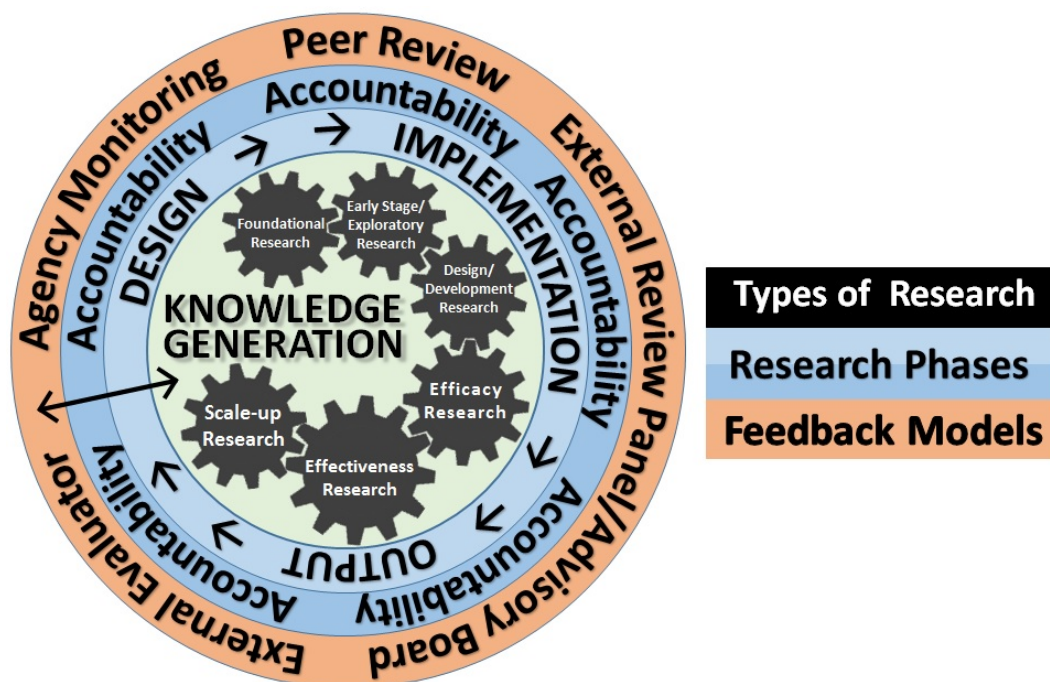
One important output from any research project is sharing the knowledge generated from the research results through various mechanisms. Research results can be packaged into an article, chapter, and/or book to be placed on websites, in journals, or included in other books or publications of specific organizations or institutions. Sometimes project materials that are developed (e.g., professional development materials, instructional materials, software, simulations, validated instruments and/or measures, etc.)—around which the research is conducted—are published by an organization or institution or provided to others for production/publication. In all of these instances, a peer review panel is one mechanism that can be used to ensure the information is conveyed accurately and in a manner consistent with the mission, policies, and practices of the organization or institution. The peer review is conducted to ensure the highest levels of accuracy and quality with the end user in mind. For most organizations and/or institutions, publishing guidelines exist and the peer review panel ensures those guidelines are consistently applied. Peer review of these materials can take place at the project level and/or organization or institution level. Individuals involved as peer reviewers at this stage can be either members of the project, external consultants hired by the project, members of organizations and/or institutions that voluntarily carry out this important role for their organization or are paid members of staff at these organizations. In all instances, individuals serving as peer reviewers have expertise, knowledge, and/or experience that are relevant to the submitted materials, or are hired by the institution or organization as consultants for specific reviews.

PUTTING IT ALL TOGETHER

As shown in Exhibit 7, at the core of the relationship between research types, phases of the research cycle, and feedback models is the generation of knowledge. Knowledge is generated through one or more of six research types (discussed previously). Typically, each research type seeks to strategically answer research questions through iterative phases which, as the model illustrates, typically occur in a cyclical fashion. Accountability is considered a research phase, but also is relevant for the research design, implementation, and output. The application of an appropriate feedback model or multiple feedback models during each phase of a research cycle can strengthen and enhance the quality of the research as a project unfolds.

An important element of this model is the bi-directionality of the components. Each component may influence the other components or be influenced by its neighbor. That is, the phases of the research are implemented based on the research types. For example, the design and implementation phases of Design and Development research are closely linked. Information generated in the research phases may inform the selection of a subsequent research type; as the *Common Guidelines* note, not all research types are independent of one another. Similarly, the phases of the research, including the accountability component, can be used to determine the most effective feedback model and vice versa; the feedback provided may inform the iterative phases of the research in a cycle of continuous improvement. Of course, this model may not hold true for every education research project, but it does provide a way to conceptualize the relationships among knowledge generation, research types, phases of research, and feedback models for evaluation of the research.

EXHIBIT 7. RELATIONSHIP OF RESEARCH TYPES, PHASES OF THE RESEARCH CYCLE, AND EXTERNAL FEEDBACK MODELS



PLANNING FOR EVALUATION OF RESEARCH

USE OF MULTIPLE EXTERNAL FEEDBACK MODELS

To fulfill the purpose of evaluation, it may be advantageous to utilize multiple types of external feedback. For example, Principal Investigators generally receive a peer review of their funded project prior to funding. Many funding agencies provide feedback and negotiate the details of the project and funding prior to the final award of the grant. In addition, many funding agencies require an evaluation plan for funding. When research is evaluated, it affords researchers the ability to correct course to improve the project based on the available evidence. Exhibit 8 demonstrates a schematic of how feedback models can be aligned to research types. This exhibit suggests that for the first three research types in the table (Foundational Research, Early Stage or Exploratory Research, along with Design and Development Research), the primary feedback models could be external review panels or advisory boards, given budget constraints. Of course, third-party evaluators could also be utilized for these three types of research as well, depending on the research agenda, project goals, and objectives. Typically, larger-budget and bigger-scale projects such as efficacy, effectiveness, or scale-up research may select a third-party evaluator to conduct a more rigorous evaluation of the research. These recommendations are denoted as primary feedback models by the large ① in the table. External review panels and third-party evaluators may indeed be complementary to one another, so they are often used together for large scale projects. Furthermore, peer review, including peer review of publications and presentations are effective feedback models, but may play a less central role to the research being conducted. While they complement the other evaluation feedback models, they are retrospective to the design and implementation of the research and thus cannot inform the continuous improvement of the research itself. Finally, agency monitoring and review are required for all research types.

EXHIBIT 8. EXTERNAL FEEDBACK MODELS ALIGNED TO RESEARCH TYPES

Research Types	External Feedback Model				
	Peer Review of the proposed project	Agency Monitoring and Review	External Review Panels or Advisory Boards	Third-party Evaluator	Peer Review of Publications and Conference Presentations
Foundational Research	⊙	⊙	①	②	③
Early Stage or Exploratory Research	⊙	⊙	①	②	③
Design and Development Research	⊙	⊙	①	②	③
Efficacy Research	⊙	⊙	②	①	③
Effectiveness Research	⊙	⊙	②	①	③
Scale-Up Research	⊙	⊙	②	①	③

Note. ① = primary feedback model; ② = secondary feedback model; ③ = tertiary feedback model; ⊙ = happens in all grant-funded research.

EVALUATION QUESTIONS THAT SUPPORT EVALUATION OF RESEARCH

Typical evaluation designs include decisions about and elaboration of evaluation questions, instrumentation, methodology, outcomes, implementation, partnerships, and reporting. Because these components have been detailed throughout methods papers and textbooks, this document focuses on organizing evaluation questions that are particularly suited to the evaluation of educational research. We focus on this component of research evaluation because a solid foundation of evaluation questions can be used to guide appropriate decisions that meet the needs of the research evaluation throughout the entire evaluation process. As shown in Exhibit 9, there are different types of questions that can be used for different phases of the research cycle and which would be appropriate for all research types.

EXHIBIT 9. SAMPLE QUESTIONS BY PHASE

Phase	Sample Questions
Pre-Funding Design ^a	<ul style="list-style-type: none"> To what extent is the proposed research based on prior research or best practices? To what extent are the research questions specific and operationalized? To what extent will the research design provide sufficient evidence to demonstrate the extent to which outcomes can be attributed to specific intervention activities? To what extent does the research and/or development team demonstrate sufficient experience and competence to carry out the research/development?
Design	<ul style="list-style-type: none"> Did the sample size afford sufficient statistical power to address the impact question(s)? To what extent is the sampling plan robust to unexpected loss of data (e.g., nonresponse, attrition, scheduling conflicts, etc.)? Are instruments/measure reliable and valid, or are there plans to validate them? Is the length of time and intensity of delivery of the intervention reasonable to be able to detect changes?
Implementation	<ul style="list-style-type: none"> To what extent did the communication structure used by the project support project goals? To what extent did the data collection and analysis deviate from the original plan? To what extent were mid-course corrections needed to improve implementation or evaluation activities? To what degree were appropriate data collected to ensure measurement of project processes, implementation and outcomes?
Output	<ul style="list-style-type: none"> Are appropriate methods of analysis provided to describe the sample and support the findings? Are potential confounds to findings identified and accounted for (i.e., attrition, nonresponse, etc.)? To what extent does the research provide sufficient evidence to demonstrate the extent to which outcomes can be attributed to specific intervention activities? To what extent has the project been able to document impacts on various participants, partners, and stakeholder groups?
Accountability	<ul style="list-style-type: none"> How did changes in the data collection and analysis plan affect the ability of the study to answer the original research questions? Did changes in staffing or timelines or other factors affect activities? To what degree did the project accomplish the goals? Was funding spent appropriate and were funds sufficiently expended?

^a Pre-funding design was briefly mentioned earlier in this paper in the Research Phases section. Because our main focus has been on research conducted post-funding, the pre-funding phase is acknowledged here as an important component of educational research for which evaluation questions might be selected.

The Appendix contains tables of additional sample evaluation questions, organized by the phase of research that can be easily adapted or modified to fit any research type. We believe that these questions can be useful in guiding the evaluation of the different research types, and provide a starting point to consider other aspects of the evaluation of research that are based on the objectives of the research.

STRENGTHENING EVALUATION PLANS IN RESEARCH PROPOSALS

In the *Common Guidelines*, the IES and the NSF provided a framework for the purpose, justification, and design of educational research in order to clarify and strengthen research proposals submitted to both agencies. Although the *Common Guidelines* indicate that all research projects should include some form of external evaluative feedback, little guidance is provided on how to structure that feedback. We have made the case here that evaluation questions may be particularly well suited to certain types of feedback for specific research types. In addition, the following suggestions may help proposers present a clear case for the strength of their external feedback plans in research proposals.

- Make a clear distinction between research activities and evaluation activities in the proposal. Accordingly, identify evaluation questions that are distinct from the research questions and that address each phase of the research (design, implementation, outputs, accountability), as appropriate.
- Clearly specify roles of researchers, implementation staff, and those contributing to evaluation feedback so that reviewers are confident that the external feedback will be sufficiently independent and objective. Describe how the (expected) qualifications of those providing feedback are of sufficient breadth to provide meaningful feedback on the research methodology, substantive content, target population, and any other salient implementation features.
- Name specific measures that will be adopted or adapted, if possible, instead of mentioning vague constructs. Include reliability and validity information about the selected measures, or describe the procedures that will be used to make a case for reliability and validity of measures, if this information is not available.
- Describe how often external feedback will be provided and in what format, to whom the feedback will be available, and to the extent possible, describe how researchers will respond to and incorporate feedback to improve the research during the project.

CONCLUSION

Our efforts in providing this document related to evaluation of research is to further support the work of evaluators as they move from impact evaluation models to feedback models for the evaluation of research. Our work has afforded us the opportunity to recognize that as more agencies are looking to further support research to inform a larger audience, there is a need for clarification regarding what research evaluation models provide various research types with feedback that informs the research design and methods, as well as providing evidence of impact for certain research types. By looking at a variety of possible evaluation questions that are appropriate at each phase of the research cycle, we hope evaluators are able to adapt and modify those questions to better fit the context and type of research they are evaluating. Finally, our hope is that this document serves as a conversation starter, a place to begin to better understand how research and evaluation are similar, but also how they are different. We believe that for prospective research projects, engaging an evaluator in the discussions early, so they can better inform the design of the research and proposal, projects will be better positioned at the outset to not only conduct research and share findings with a larger audience, but to use evaluation as a way to enhance the research and impacts.

REFERENCES AND ARTICLES REVIEWED TO DEVELOP MODEL FOR EVALUATION OF RESEARCH

- Abell, S. K., & Lederman, N. G. (2013). *Handbook of Research on Science Education*. New York: Routledge.
- Arthur, L., & Cox, E. (2014). From evaluation to research. *International Journal of Research & Method in Education*, 37:2, 137–150. doi:10.1080/1743727X.2013.795529.
- Baizerman, M. L., Fink, A., & VeLure Roholt, R. (2012). From consilium to advice: A review of the evaluation and related literature on advisory structures and processes. In R. VeLure Roholt & M. L. Baizerman (Eds.), *Evaluation advisory groups. New Directions for Evaluation*, 136, 5–29.
- Brutscher, P. B., Wooding, S., & Grant, J. (2008). *Health Research Evaluation Frameworks: An International Comparison*. Technical Report. Cambridge: RAND Corporation.
- Cahalan, M., & Goodwin, D. (2014). Setting the record straight: Strong positive impacts found from the national evaluation of Upward Bound. Washington, DC: The Pell Institute for the Study of Opportunity in Higher Education.
- Callow-Heusser, C., Chapman, H. J., & Torres, R.T. (2005). *Evidence: An essential tool: Planning for and gathering evidence using the design-implementation-outcomes (DIO) cycle of evidence*. National Science Foundation Math and Science Partnership Program.
- Carlson, E., Bitterman, A., Zhang, X., Gutmann, B., Wills, K., & Sinclair, B. (2015). *Evaluation of the Regional Educational Laboratories: Final Report* (NCEE-2015-4008). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Cohen, B. B. (2012). Advisory groups for evaluations in diverse cultural groups, communities, and contexts. In R. VeLure Roholt & M. L. Baizerman (Eds.), *Evaluation advisory groups. New Directions for Evaluation*, 136, 49–65.
- Confrey, J. (2006). Comparing and contrasting the National Research Council Report “On Evaluating Curricular Effectiveness” with the What Works Clearinghouse Approach. *Educational Evaluation and Policy Analysis*, 28, (3) 195–213.
- Derrick, G. E., & Pavone, V. (2013). Democratising research evaluation: Achieving greater public engagement with bibliometrics-informed peer review. *Science and Public Policy*, 40, 563–575. doi:10.1093/scipol/sct007.
- Desimone, L. M. (2009). Improving impact studies of teachers’ professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38, 181–199.

- Doabler, C. T., Clarke, B., Fien, H., Baker, S. K., Kosty, D. B., & Cary, M.S. (2015). The science behind curriculum development and evaluation: Taking a design science approach in the production of a Tier 2 mathematics curriculum. *Learning Disability Quarterly*, 38, 97–111.
- Ellet, C.D., & Monsaas, J. (2011). Development and validation of a new measure of teacher perceptions of science and mathematics learning environments. *Learning Environment Resources*, 14, 91–107.
- Ferrer-Wreder, L., Adamson, L., Kumpfer, K. L., & Eichase, K. (2012). Advancing intervention science through effectiveness research: A global perspective. *Child Youth Care Forum*, 41:109–117.
- Frechtling, J. (2010). The 2010 User-Friendly Handbook for Project Evaluation. Division of Research and Learning in Formal and Informal Settings. Arlington, VA: National Science Foundation.
- Fu, A. C., Peterson, L., Kannan, A., Savelson, R. J., & Kurpius, A. (2015). A framework for summative evaluation in informal science education. *Visitor Studies*, 18, 17–38.
- Gast, David L. (2014). *Single case research methodology: Applications in special education and behavioral sciences (2nd ed.)*. Gast, David L., & Ledford, Jennifer R. (Eds). New York, NY: Routledge/Taylor & Francis Group; 2014, pp. 85–104.
- Gast, David L.; Lloyd, Blair P.; Ledford, Jennifer R. (2014). *Single case research methodology: Applications in special education and behavioral sciences (2nd ed.)*. Gast, David L., & Ledford, Jennifer R. (Eds). New York, NY: Routledge/Taylor & Francis Group; 2014, pp. 251–296.
- Gersten, R. & Hitchcock, J. (2009). What is credible evidence in education? The role of the What Works Clearinghouse in informing the process. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 78-95). Los Angeles, CA: SAGE.
- Grant, J., Brutscher, P. B., Kirk, S. E., Butler, L., & Wooding, S. (2010). Capturing research impacts: A review of international practice. Santa Monica, CA: RAND Corporation.
- Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S., & Grant, J. (2013). *Measuring research: A guide to research evaluation frameworks and tools*. Santa Monica, CA: RAND Corporation.
- Hage, J., Jordan, G., & Mote, J. (2007). A theory-based innovation systems framework for evaluating diverse portfolios of research, part two: macro indicators and policy interventions. *Science & Public Policy*. 34(10), 731–741.
- Hallinger, P. (2013). A conceptual framework for systematic reviews of research in educational leadership and management. *Journal of Educational Administration*, 51, 126–149. doi:10.1108/09578231311304670.
- Institute of Education Sciences, U.S. Department of Education, and the National Science Foundation. (2013, August). *Common Guidelines for Educational Research and Development*. Washington, DC. Available at <http://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf>.

- Jacob, B. & Ludwig, J. (2005). Can the federal government improve education research? *Brookings Papers on Education Policy*, 8, 47–87.
- Katzenmeyer, C. & Lawrenz, F. (2006). National Science Foundation perspectives on the nature of STEM program evaluation. *New Directions for Evaluation*, 109, 7–18.
- Kelley, A. E., Lesh, R. A., & Baek, J. Y. (Eds.) (2008). *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*. New York, NY: Routledge/Taylor & Francis.
- Kinzie, M. B., Whittaker, J. V., McGuire, P., Lee, Y., & Kilday, C. (2015). Research on curricular development for pre-kindergarten mathematics and science. *Teachers College Record*, 117, 1–40.
- Knowles, T., Kober, N., Ludwig, J., Petrilli, M. J., Slavin, R., & Tseng, V. (2013). *Leveraging learning: The evolving role of federal policy in education research*. The Aspen Institute: Washington, D.C.
- Kwan, B. S. C., Chan, H., & Lam, C. (2012). Evaluating prior scholarship in literature reviews of research articles: A comparative study of practices in two research paradigms. *English for Specific Purposes*, 31, 188–201. doi:10.1016/j.esp.2012.02.03.
- Lawrenz, F. (2007). Review of science education program evaluation. In S. K., Abell, & N.G., Lederman (Eds.), *Handbook of Research on Science Education* (pp. 943–963). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Leech, N. L., Dellinger, A. B., Brannagan, K. B., & Tanaka, H. (Eds.) (2010). Evaluating mixed research studies: A mixed method approach. *Journal of Mixed Methods Research*, 4(1), 17–31.
- Madaus, G. F., Scriven, M. S. & Stufflebeam, D. L. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer-Nijhoff Publishing.
- Marjanovic, S., Hanney, S., Wooding, S. (2009). *An historical reflection on research evaluation studies, their recurrent themes and challenges*. Santa Monica, CA: RAND Corporation.
- Mateo, J., Escofet, A., Martinez-Olmo, F., Ventura, J., & Vlachopoulos, D. (2012). Evaluation tools in the European higher education area (EHEA): An assessment for evaluating the competencies of the final year project in the social sciences. *European Journal of Education*, 47(3), 435–447.
- Mattessich, P. W. (2012). Advisory committees in contract and grant-funded evaluation projects. In R. VeLure Roholt, & M. L. Baizerman (Eds.), *Evaluation advisory groups*. *New Directions for Evaluation*, 136, 31–48.
- Milesi, C., Brown, K. L., Hawkley, L., Dropkin, E., & Schneider, B. L. (2014). Charting the impact of federal spending for education: A bibliometric approach. *Educational Researcher*, 43(7), 361–370.
- Murnane, R. J., & Willett, J. B. (2011). *Methods Matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

- Newton, X. A., Poon, R. C., Nunes, N. L., & Stone, E. M. (2013). Research on teacher education programs: Logic model approach. *Evaluation and Program Planning*, 36, 88-96.
- Oancea, A. (2007). From procrestes to Proteus: Trends and practices in the assessment of education research. *International Journal of Research & Method in Education*, 30(3), p243–269.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational research*, March 1, 2008 78:33-84.
- Research Excellence Framework. (2012). *Research Excellence Framework*. Retrieved from www.ref.ac.uk/
- Richey, R. C., & Klein, J. D. (2007). Design and development research. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Sample McMeeking, L. B., Basile, C. B., & Cobb, R. B. (2012). An organizational model to distinguish between and integrate research and evaluation activities in a theory based evaluation. *Evaluation and Program Planning*, 35, 508–516.
- Scherer, J. (2008). A review of instruments to evaluate partnerships in math and science education. *Peabody Journal of Education*, 83, 611–636.
- Scriven, M. & Coryn, C. (2008, Summer). The logic of research evaluation. *New Directions for Evaluation*, 118, 89–105.
- Stecher, B., Le, V., Hamilton, L., Ryan, G., Robyn, A., & Lockwood, J.R. (2006). Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educational Evaluation and Policy Analysis*. Vol 28 (2), Sum 2006, p101–130. Washington, DC: American Educational Research Association.
- Stige, B., Malterud, K., & Midtgarden, T. (2009). Toward an agenda for evaluation of qualitative research. *Qualitative Health Research*, 19, 1504–1516.
- Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., Unger, J., Palmer, P., Clark, M., Colby, S., Morgan, G., & Trochim, W. (2003). *Evaluating transdisciplinary science*. Bethesda, MD: National Institutes of Health.
- Stufflebeam, D. L. Madaus, G. F, Kellaghan, T. (2000). *Evaluation Models: Viewpoints on educational and human service evaluation*. Boston, MA: Kluwer-Nijoff Publishing.
- Stufflebeam, D. L. & Webster, W. (1980, May-June). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis*. 3,(2), 5–9. Washington, DC: American Educational Research Association.
- Sutter, L. E. & Frechtling, J. (2000). *Guiding principles for mathematics and science education research methods: Report of a workshop*. Arlington, VA: National Science Foundation.

- Tash, W. (2006). *Evaluating research centers and institutes for success!: A manual and guide with case studies*. Bethesda, MA: National Institutes of Health.
- Trochim, W. M., Marcus, S. E., Mâsse, L. C., Moser, R. P., & Weld, P. C. (2008). The evaluation of large research initiatives: A participatory integrative mixed-methods approach. *American Journal of Evaluation*, 29, 8–28.
- Urban, J. B., & Trochim, W. (2009). The role of evaluation in research-practice Integration: Working toward the “Golden Spike.” *American Journal of Evaluation*, 30(4), 538–553.
- VeLure Roholt, R., & Baizerman, M. L., (2012). A model for evaluation advisory groups: Ethos, professional craft knowledge, practices, and skills. In R. VeLure Roholt, & M. L. Baizerman (Eds.), *Evaluation advisory groups. New Directions for Evaluation*, 136, 119–127.
- Wooding, S., Nason, E., Klautzer, L., Rubin, J., Hanney, S., & Grant, J. (2007). *Policy and practice impacts of research funded by the Economic and Social Research Council: A case study of the Future of Work programme, approach and analysis*. Santa Monica, CA: RAND Corporation.
- Yin, R. K., Hackett, E. J., & Chubin, D. E. (2008). Discovering “What’s Innovative”: The Challenge of Evaluating Education Research and Development Efforts. *Peabody Journal of Education*, 83(4), 2008 Special Issue: Math and science partnership program: A first comprehensive evaluation. New York, NY: Routledge/Taylor & Francis Group; 2008, pp. 674-690.

APPENDIX

This section contains tables of additional sample evaluation questions, organized by the phase of research that can be easily adapted or modified to fit any research type. We believe that these questions can be useful in guiding the evaluation of the different research types, and provide a starting point to consider other aspects of the evaluation of research that are based on the objectives of the research.

EXHIBIT A1. PRE-FUNDING AND DESIGN EVALUATION QUESTIONS

Sample Questions

Prior Research

- To what extent is the proposed research based on prior research or best practices?
- To what extent do studies build on research and theory to examine the associations between (a) education or learning outcomes and (b) malleable factors (that is, factors that are alterable, such as children’s behaviors; technologies; education programs, policies, and practices)? (Common Guidelines)
- To what extent do studies build on research and theory to examine the factors and conditions that may mediate or moderate the relationship between (a) education or learning outcomes and (b) malleable factors? (Common Guidelines)
- To what extent do studies build on research and theory to examine the opportunities for new interventions or strategies, and challenges to their adoption, with the goal of informing policy, practice, and future design or development? (Common Guidelines)

Intervention/Treatment

- To what extent did the project consider the appropriate content standards in developing the instructional modules?
- In what ways has the project considered student engagement in the instructional module development?
- Is the length of time and intensity of delivery of the intervention reasonable to be able to detect changes?
- Is the design and development of the intervention theoretically, empirically, and practically grounded in evidence? (Fu, Peterson, & Kannan, 2015)
- To what extent is the educational content of the research relevant to the research goal?

Research Questions

- To what extent are the research questions explicitly stated, aligned with the study, and empirically testable? (Carlson et al., 2015)
- To what extent are the research questions specific and operationalized?

Research Design

- Is the research design supported by relevant research or theory? (Callow-Heusser, Chapman, & Torres, 2005).
 - To what extent is the research design valid? (Callow-Heusser, Chapman, & Torres, 2005).
 - To what extent will the research design provide sufficient evidence to demonstrate the extent to which outcomes can be attributed to specific intervention activities?
-

Sample Questions

- To what extent does the research support the development of a well-explicated theory of action that can inform the development, modification, or evaluation of an intervention or strategy? (Common Guidelines)
- To what extent is the reasoning of the research explicitly, clearly, and soundly reflected in the research design?
- To what extent will the research design provide sufficient evidence to answer the research questions?
- To what extent do the research activities answer the key research questions?
- For studies comparing multiple groups, are methods of establishing comparability of groups built into the design? (Confrey, 2006)
- To what extent has the nature of the mixing been described explicitly (for mixed-methods designs)?
- To what extent have the needs of different stakeholder groups been considered in the design of the research? (e.g., to what extent have the needs of underrepresented groups been considered?) (Callow-Heusser, Chapman, & Torres, 2005)
- How has the research incorporated issues of social equity into the research design?

Context

- To what extent does the research design plan for the documentation of the relevant features of the context of the project?
- What changes in the context could affect the feasibility of the design? (Callow-Heusser & Chapman, & Torres 2005)
- How will the evidence generated support replication of the research to achieve similar outcomes in other contexts? (Callow-Heusser, Chapman, & Torres, 2005)

Fidelity

- To what extent does the research plan include measurement of the fidelity of intervention implementation?
- To what extent is the plan to document or the documentation of implementation of the intervention (e.g., exposure, quality of delivery, adherence) appropriate? (Adapted from Carlson et al., 2015)

Instruments/Measures

- What are different ways the project could/should measure growth around the important construct of “content and pedagogical content knowledge?”
- To what extent are quantitative data collection instruments valid, reliable, and objective?
- If instruments do not have reliability and validity data or are locally developed, are there plans to validate them?
- To what extent are qualitative data collection protocols credible, dependable, and confirmable?

Data/Evidence

- To what extent is the data collection plan appropriate for the research questions? (Carlson et al., 2015)
- What evidence demonstrates whether or not anticipated research outputs were achieved? (Callow-Heusser, Chapman, & Torres, 2005)
- To what extent is the planned evidence sufficient to answer the research question(s)?

Population/Sample

- To what extent is the educational content of the research appropriate for the target population?
-

Sample Questions

- To what extent is the educational content of the research relevant to the population being studied?
- To what extent were the sampling strategies appropriate for the targeted populations? (Carlson et al., 2015)
- Did the sample size afford sufficient statistical power to address the impact question(s)? (Carlson et al., 2015)
- To what extent is the sampling plan robust to unexpected loss of data (e.g., nonresponse, attrition, scheduling conflicts, etc.)?
- To what extent is an infrastructure in place to recruit participants?

Data Analysis

- Is the data analysis appropriate for answering the research questions?
- Is the data analysis appropriate for the research design?

Reporting

- Is there an appropriate plan for reporting/dissemination?

Logistics

- Can activities based on the project design be accomplished in the given timeframe with the given resources within the given context? (Callow-Heusser, Chapman, & Torres, 2005)
 - To what extent are budgets, timelines, and staffing reasonable?
 - To what extent does the research and/or development team demonstrate sufficient experience and competence to carry out the research/development and data analysis?
-

EXHIBIT A2. IMPLEMENTATION EVALUATION QUESTIONS

Sample Questions

Leadership

- To what extent does the project leadership effectively support progress toward research goals?
- To what extent did the communication structure used by the project support project goals?
- To what extent are research partners (e.g., school staff, intervention staff, participants, etc.) satisfied with the research process?

Needs

- What is the process for deciding among the many needs of the targeted audiences?
- To what extent do needs gathered by various processes inform the activities of the research?

Implementation

- To what extent was the participant recruitment plan followed with fidelity?
- To what extent was the random assignment rigorous? (Carlson et al., 2015)
- Were appropriate data collected to ensure measurement of project processes, implementation and outcomes?
- How well is the data collection plan implemented? (Carlson et al., 2015)
- To what extent did the data collection and analysis deviate from the original plan?

Challenges/Changes

- Were mid-course corrections needed to improve implementation or evaluation activities? Were reasonable adjustments made? Were challenges or barriers identified/addressed?
- What challenges has the project encountered in conducting the research?
- How effectively has the research team responded to unexpected challenges?
- How has the research team responded to unexpected changes in context over the course of the study, if any?
- How has the research team responded to unexpected threats to the successful completion of the project?

Dissemination

- To what extent is the process for selecting the delivery mechanisms for the dissemination appropriate?
- To what extent is the project team effectively carrying out the dissemination plan?

Logistics

- To what extent did the research follow the planned timeline?
 - To what extent did the research follow the original budget?
-

EXHIBIT A3. OUTPUTS EVALUATION QUESTIONS

Sample Questions

Analysis

- Are all research questions specifically addressed by analyses? (Carlson et al., 2015)
- Are appropriate methods of analysis provided to describe the sample and support the findings? (Carlson et al., 2015)
- To what extent does the research provide sufficient evidence to demonstrate the extent to which outcomes can be attributed to specific intervention activities? (Callow-Heusser, Chapman, & Torres, 2005)
- Are potential confounds to findings identified and accounted for (i.e., attrition, nonresponse, etc.)?
- To what extent has the project been able to document impacts on various participants, partners, and stakeholder groups?
- Are conclusions drawn appropriately and consistently? (Carlson et al., 2015)

Reporting

- What were the technical quality and relevance of study reports produced by the research and of the corresponding proposals? (Carlson et al., 2015)
- To what extent do reports provide a thorough summary of key literature and/or previous research in this topic area? (adapted from Carlson et al., 2015)
- To what extent were the intervention for the treatment group and the condition for the control group clearly described? (adapted from Carlson et al., 2015)
- To what extents were results reported disaggregated by subgroups?
- To what extent are limitations of the study clearly and comprehensively stated? (Carlson et al., 2015)
- To what extent do reports contribute to new information about the relationships between predictors and outcomes being studied? (adapted from Carlson et al., 2015)
- How are the research processes and products useful for practice and understanding in relation to real world problems and situations? (Stige, Malterud, & Midtgarden 2009)
- To what extent is the report format appropriate for the intended audience?

Dissemination

- To what extent have research outputs reached the intended audience?
 - What are the strategies being considered to share what was learned with the larger education community?
 - How did researchers and stakeholders use research findings and reports?
 - To what extent are research outputs relevant to practitioners in addition to researchers?
 - To what extent do research outputs reach both researchers and practitioners?
 - In what ways has the project influenced or informed the evaluation of other projects?
 - To what extent are data gathered in the research available for further analysis by other researchers?
 - To what extent does the research support the replication of similar outcomes in other contexts? (Callow-Heusser, Chapman, & Torres, 2005)
 - What factors influence the diffusion, adoption, and adaptation of the innovation or intervention? (adapted from Kelley, Lesh, & Baek, 2008)
-

EXHIBIT A4. ACCOUNTABILITY EVALUATION QUESTIONS

Sample Questions

Goals

- To what degree did the project accomplish the goals?
- To what extent did the research project achieve more than originally intended?
- If the project did not accomplish its goals, to what extent will the knowledge that was produced contribute to the knowledge of the field?
- To what extent does the research contribute to new information about the relationships between predictors and outcomes being studied? (adapted from Carlson et al., 2015)

Changes

- To what extent has the research team been responsive to feedback?
- What changes, if any, have been made to the research plan to accommodate changes in the design, data collection, and implementation process?
- How did changes in the data collection and analysis plan affect the ability of the study to answer the original research questions?
- Did changes in staffing or timelines or other factors affect activities?

Logistics

- To what extent was the research carried out with reasonable timeline and sequencing of activities?
- What evidence demonstrated project/activity goals were reached on time and within budget?
- To what extent were funding resources used efficiently and effectively?
- Was funding spend appropriately and were funds sufficiently expended??

Scaling Up

- To what extent has progress been made towards sustaining or “scaling up” the intervention?
-

EXHIBIT A5. COMMON GUIDELINES EVALUATION QUESTIONS

Sample Questions

Intervention

- To what extent did the researchers develop a solution (for example, an instructional approach; design and learning objects, such as museum exhibits or media; or education policy) based on a well-specified theory of action appropriate to a well-defined end user?
- To what extent did the researchers develop new or improved interventions or strategies to achieve well-specified learning goals or objectives?
- To what extent did the researchers investigate approaches to education problems and establish the basis for design and development of new interventions or strategies?
- To what extent did the researchers create measures to assess the implementation of the intervention?

Research Type

- Did the researchers conduct a pilot study to examine the promise of generating the intended outcomes?
- To what extent did the research provide evidence for whether an established intervention or strategy is ready to be tested in an efficacy study?
- To what extent did the research support the determination of whether an intervention or strategy can improve outcomes under ideal conditions?
- To what extent did the research estimate the impacts of an intervention or strategy when implemented under conditions of routine practice?
- Did the researchers collect data on the feasibility of implementing the solution/intervention in typical delivery settings by intended users?
- To what extent was the research able to estimate the impacts of an intervention or strategy under conditions of routine practice and across a broad spectrum of populations and settings?

Findings/Broader Impact

- To what extent did the research provide fundamental knowledge about teaching and/or learning?
 - To what extent did the research develop and refine theory and methodology?
 - To what extent did the research advance the frontiers of education and learning?
 - How did researchers and stakeholders use research findings and reports?
-